MIE09 SARAJEVO
Medical Informatics in a United and Healthy Europe
30 August to 2 September 2009
www.mie2009.org

The Role of Ontologies for Sustainable, Semantically Interoperable and Trustworthy EHR Solutions

Semantic enrichment of narrative EHR content

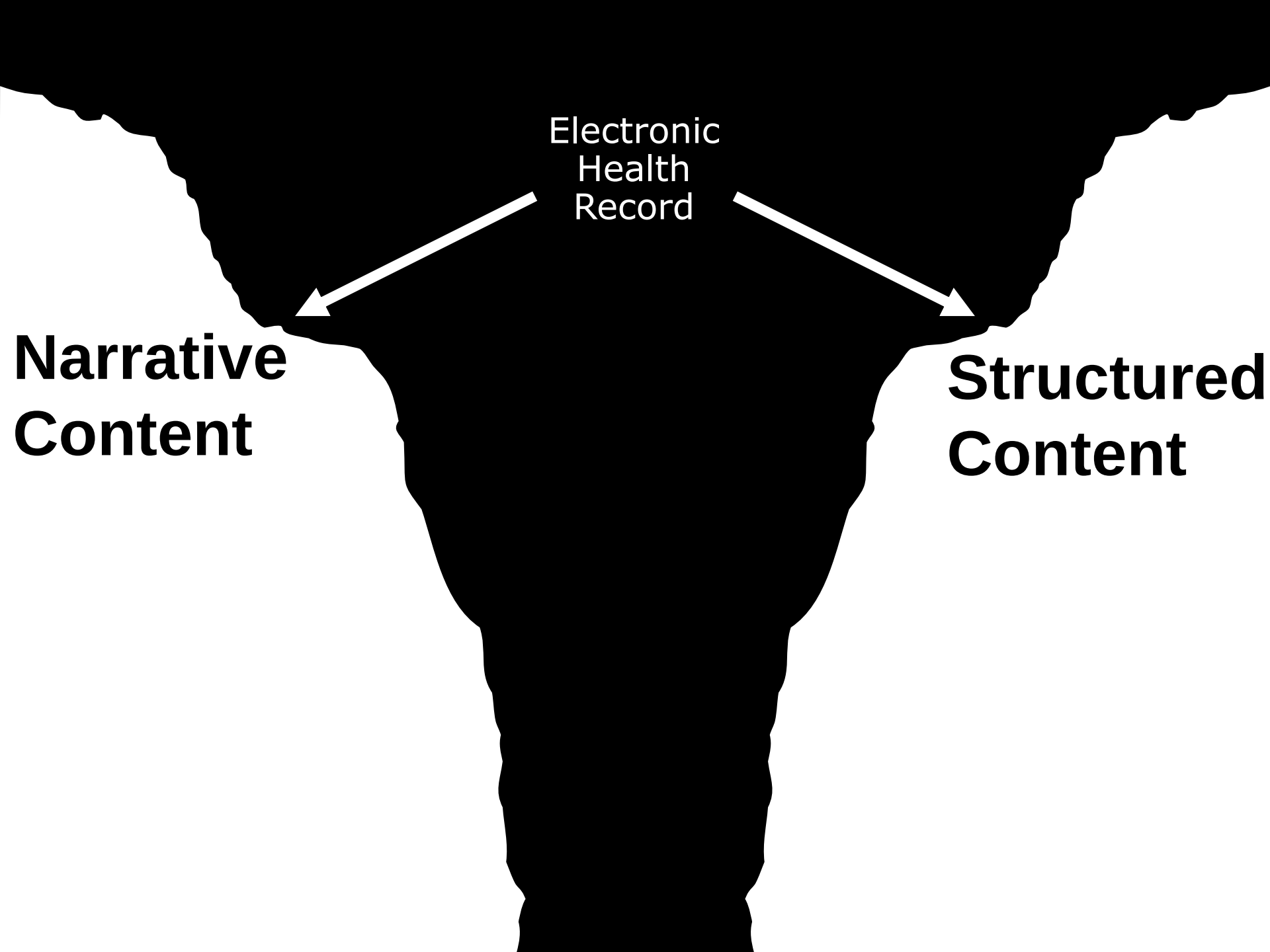
Stefan SCHULZ

Institute of Medical Biometry and Medical Informatics

University Medical Center, Freiburg, Germany

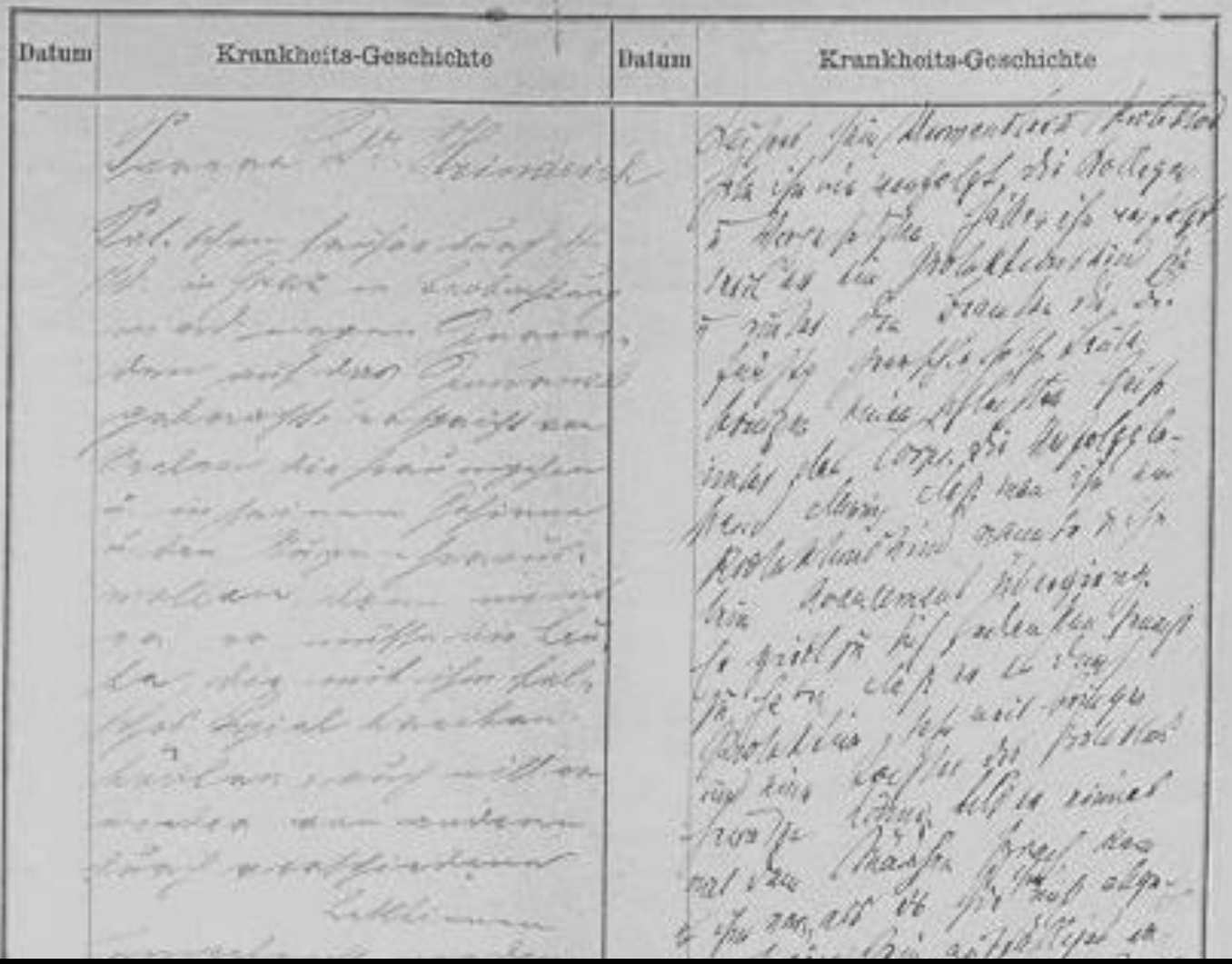
Structure of the talk

- The role of natural language in the electronic health record
  - different types of narratives and different ways of authoring
  - advantages and disadvantages of narrative content
  - need of semantic enrichment of the EHR
- Target representations for semantically enriched EHRs
  - ontologies are not enough
  - context is essential
- Language engineering for semantic EHR enrichment: Technical challenges


Electronic
Health
Record
Narrative
Content
Structured
Content

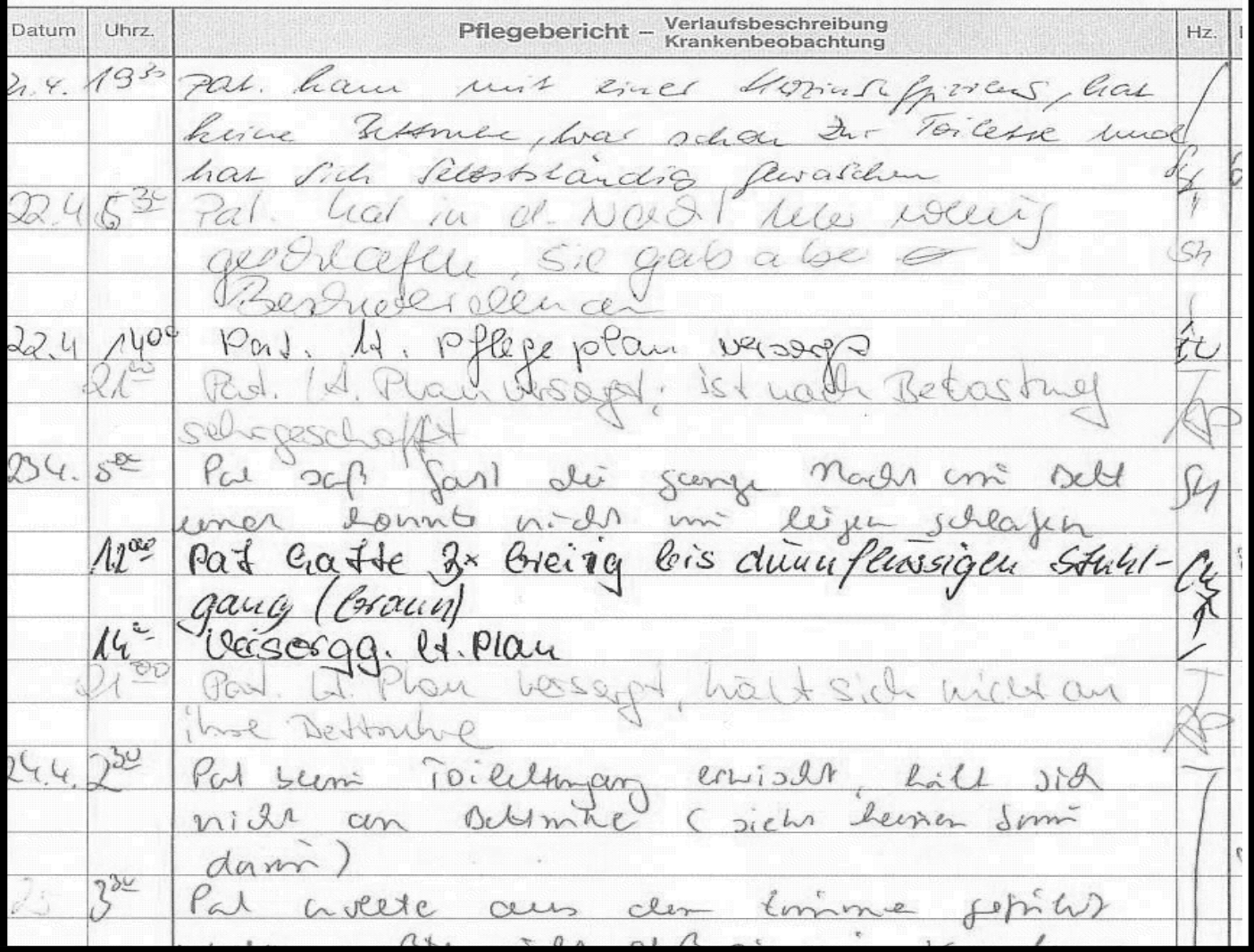
The importance of narrative EHR content

Handwritten notes (clinical history)



Source: S. Freud's handwritten clinical notes

Handwritten notes (nursing report)



Datum	Uhrz.	Pflegebericht – Verlaufsbeschreibung Krankenbeobachtung	Hz.
21.4.	19[30]	Pat. kam mit einer Harnwegsinfekt, hat keine Schmerzen, war schon zur Toilette und hat sich selbstständig gewaschen	
22.4	5[30]	Pat. hat in d. Nacht gut geschlafen, sie gab ab. Beschwerden	
22.4	14[00]	Pat. h. Pflegeplan besorgt	
	21[00]	Pat. h. Plan besorgt; ist nach Betastung schlafgeschafft	
23.4.	5[00]	Pat sah fast die ganze Nacht und hat immer kaum nicht im Liegen schlafen	
	11[00]	Pat hatte 3× breiig bis dünnflüssigen Stuhlgang (braun)	
	14[00]	besorgt h. Plan	
	21[00]	Pat. h. Plan besorgt, hält sich nicht an ihre Bettruhe	
24.4.	2[30]	Pat kann Toilettengang erwidert, hält sich nicht an Bettruhe (sieht besser dann)	
	3[30]	Pat wollte aus dem Zimmer gehen	

Source: Universitätsklinikum Freiburg, Germany

Discharge Summary dictated by MD, written by professional typist

Vater verstorben an Bronchial-Karzinom, Mutter verstorben an den Folgen einer Pneumonie. Mutter Diabetes mellitus. 5 gesunde Kinder.

Systemanamnese:

Derzeit Appetitlosigkeit, Trockengewicht um 75 Kg, derzeit 80 Kg. Miktio: gelegentlich Harn-verhalt, gehäuft Harnwegsinfekte, derzeit keine Algurie. Vor Dialyse keine Rest-Diurese. Stuhlgang obstipiert, benutzt regelmäßig Abführmittel. Vor NTX starker Juckreiz, Seit NTX deutlich rückläufig. Kein Husten/Auswurf. Noxen: Nichtraucherin, kein Alkohol.

Soziale Anamnese:

Früher Arbeiterin in der Elektronikbrache, dann Hausfrau, verheiratet, lebt mit dem Ehemann zusammen.

Allergien. Keine bekannt.

Medikation bei Aufnahme:

Ulcogant 1-1-1, Pepdul mit 0-0-0-1, Cellcept 2x1 g, Bayotensin 3 x 1, Cynt 0,2 1x1, Ludiomil 50 mg 1 x 1, Sandimmun 2 x 150 mg, Clexane 0,4 ml 1 x täglich s.c.

Status bei Übenahme:

58-jährige Patientin in vorgealtertem, reduziertem Allgemein- und adipösem Ernährungszu-stand (80 Kg Gewicht bei 160 cm Körpergröße). RR 170/80 mm Hg, Puls 66/Minute, regelmäßig. Punktförmige Depigmentierungen an beiden Unterarmen bei Zustand nach heftigem Kratzen wegen Juckreiz. Keine zervikalen Lymphome. Mundschleimhaut trocken, Zunge weißlich belegt. Rachenschleimhaut reizlos, Tonsillen schlecht einsehbar. Schilddrüse nicht vergrößert. Pulmo: Sonorer Klopfschall und vesikuläres Atemgeräusch. Cor: Spitzenstoß nicht tastbar, leise, reine Herztöne. 3/6. spindelförmiges Systolikum und 1-2/6. Decrescendo-Sofort-dialstolikum über der Aorta mit Fortleitung in die Karotis. Kein abdominales und inguinales Strömungsgeräusch. Abdomen: Bei Adipositas Organgrenzen schlecht beurteilbar, Leber/Milz nicht vergrößert. Reizlose Narbe im Bereich des rechten Unterbauches bei Zustand nach NTX. Dort leichte Druckdolenz. Wirbelsäule nicht klopfschmerzhaft. Bds. Unterschen-kelödeme. Feinschlägiger Tremor beim Arm-Vorhalte-Versuch. Pupillen isokor, Lichtreaktion prompt. Finger-Nase-Versuch bds. unsicher, ataktisch. Reflexe seitengleich.

Source:

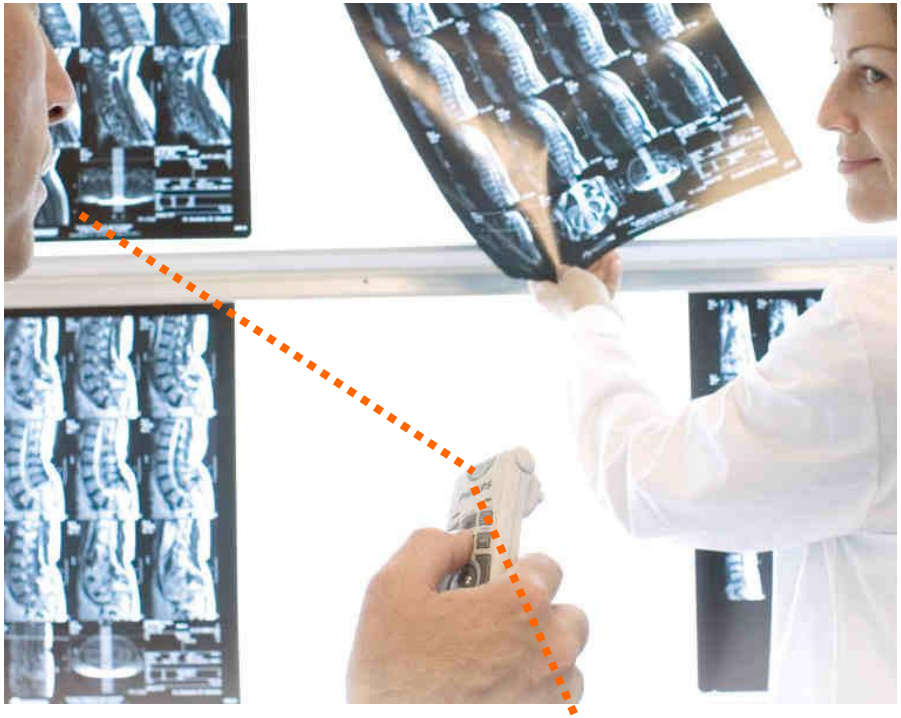
Source: Universitätsklinikum Freiburg, Germany

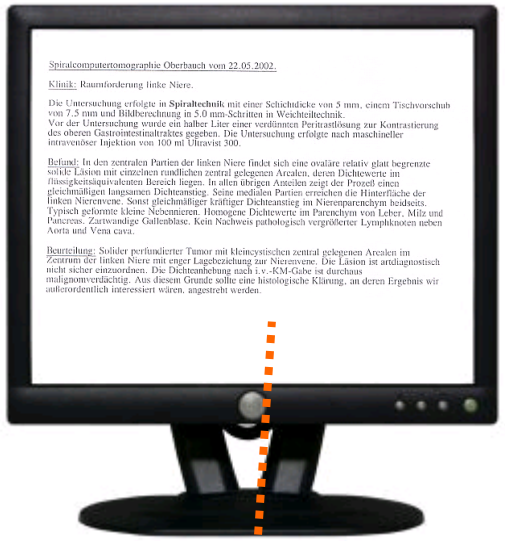
Discharge summary, typed by MD, not proofread

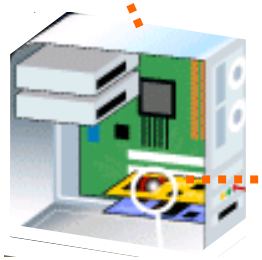
Erotilde, 58 anos # HAS # Obesidade Pcte interna com hist de edema e dor em MID há 3 semanas. Refere que no inicio do quadro apresnetava hiperemia local importante, nega febre. Foi avaliada no posto de saude e iniciado amoxacilina. Fez uso do ATB por 6 dias, com piora da dor, do edema, e surgimento de lesões arrendondadas, planas, com bordos bem definido e pequeno ponto escurecido central. Consultou novamente no posto de saude, snedo trocado o ATB para eritromicina, o qual usou por mais 6 dias, com piora das lesões e da dor. Na avaliação inical a pcte apresentava edema e hiperemia imporante de MID, bem como lesoes ulceradas, necróticas com bordos bem definidos, sem secreção. Foi realizado ECO doppler que confirmou TVP em MID sendo então iniciada anticoagulação com enoxaheparina. Solicitada consultoria da dermato que realizou biopsia das lesões
AP: vasculite leucocitoclastica. A pcte recebeu ciprofloxacin por 5 dias e após 2 dias de oxacilina. Recebe alta em bom estado geral, com diminuição importante do edema e da dor em MID. Lesões em fase de cicatrização. Revisada a literatura: existe associação de vasculite em areas de estase, bem como associação com farmacodermia. Em uso de: Varfarin 5mg 1cp por dia Captopril 25mg TID Amitrptilina 50mg Fluoxetina 40mg HCTZ 25 mg Paciente submetida a fundoplicatura videolaparoscópica. Recebe alta aceitando bem a via oral, sinais vitais estáveis, sem intercorrências ao longo da internação. Plano de retorno ambulatorial.

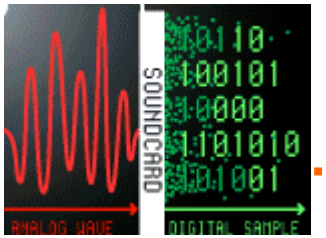
Source: Hospital de Clínicas de Porto Alegre, Brazil

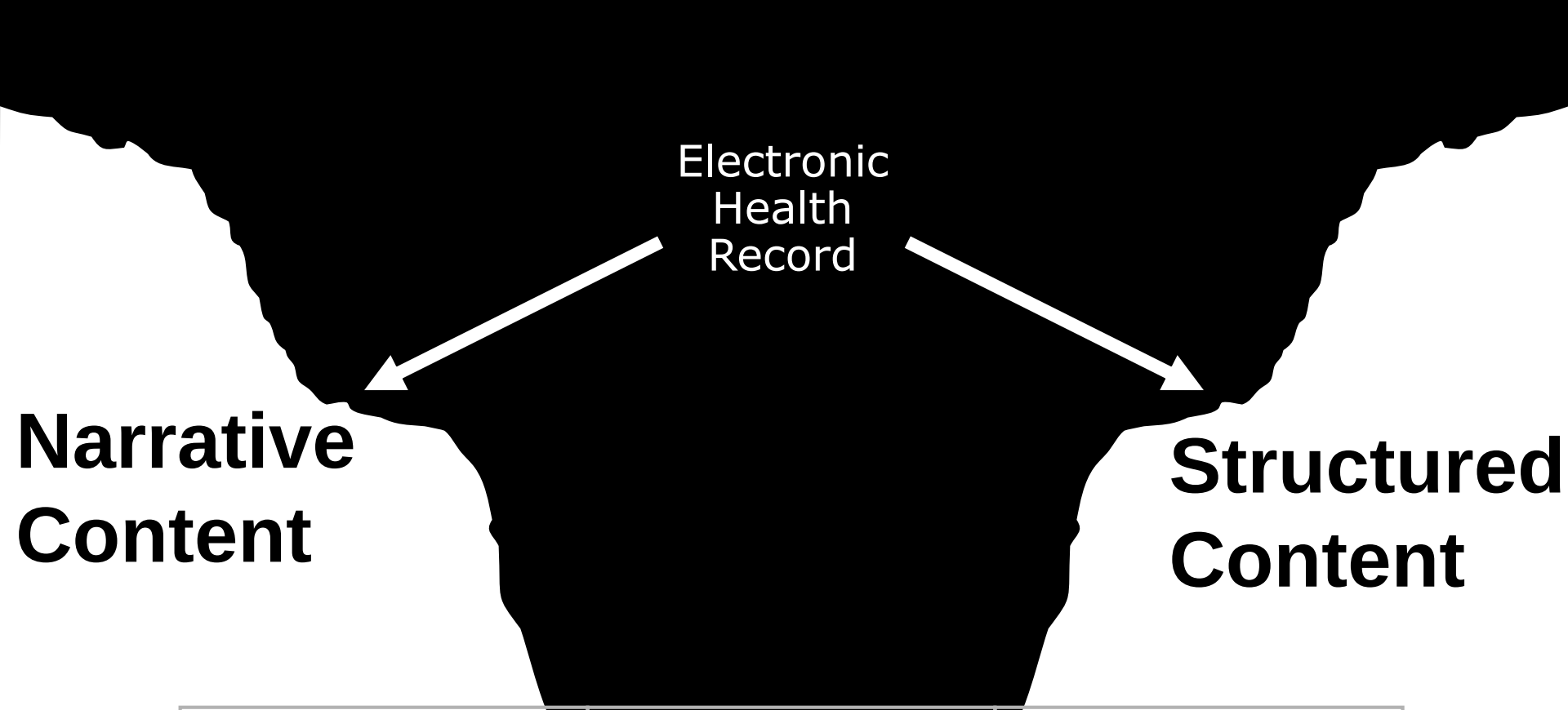
Narratives produced by speech recognition





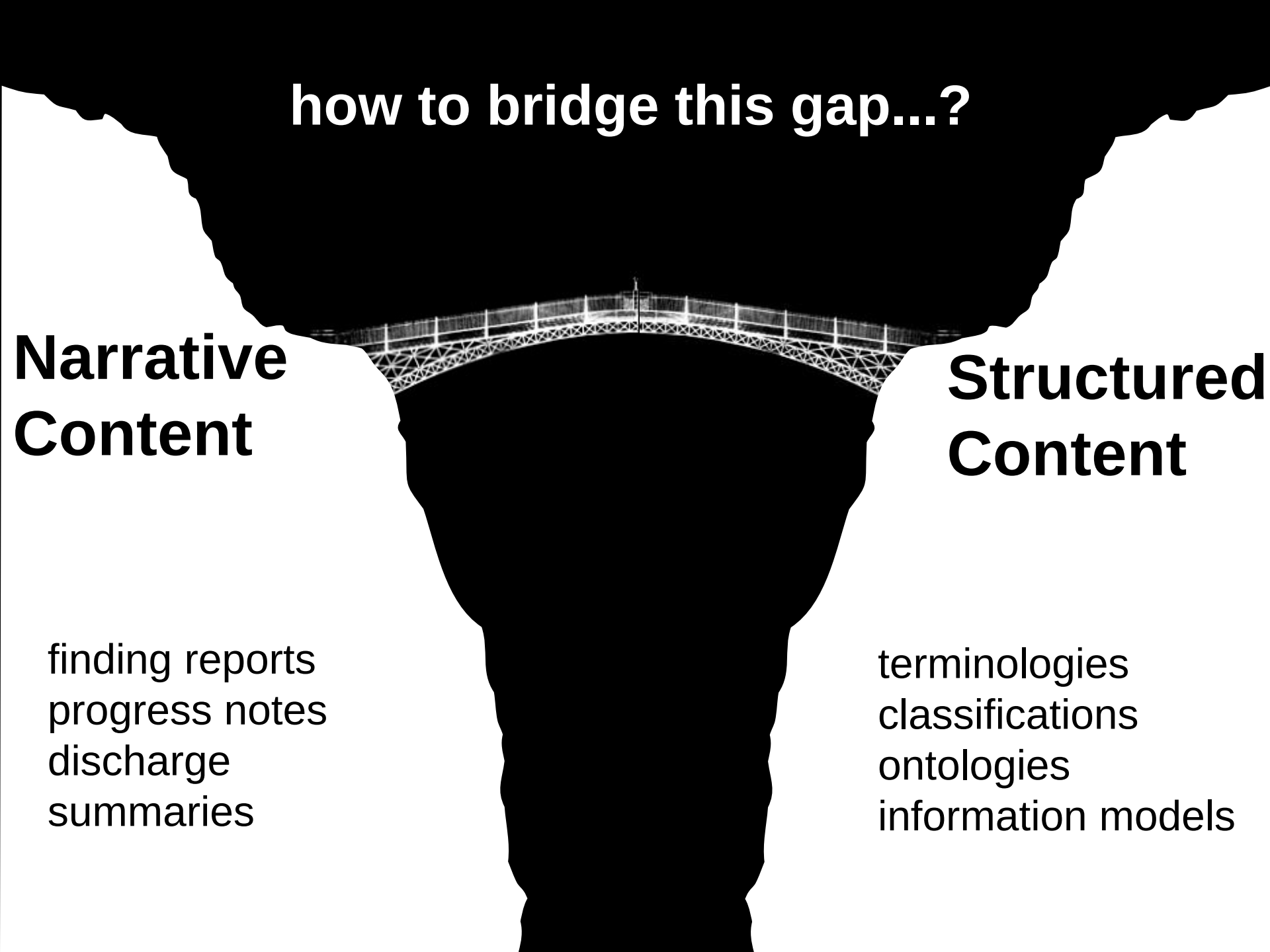






+	Ease of production	-
+	Quality	-
+	human communication	-
-	machine communication	+
-	data analysis	+

how to bridge this gap...?



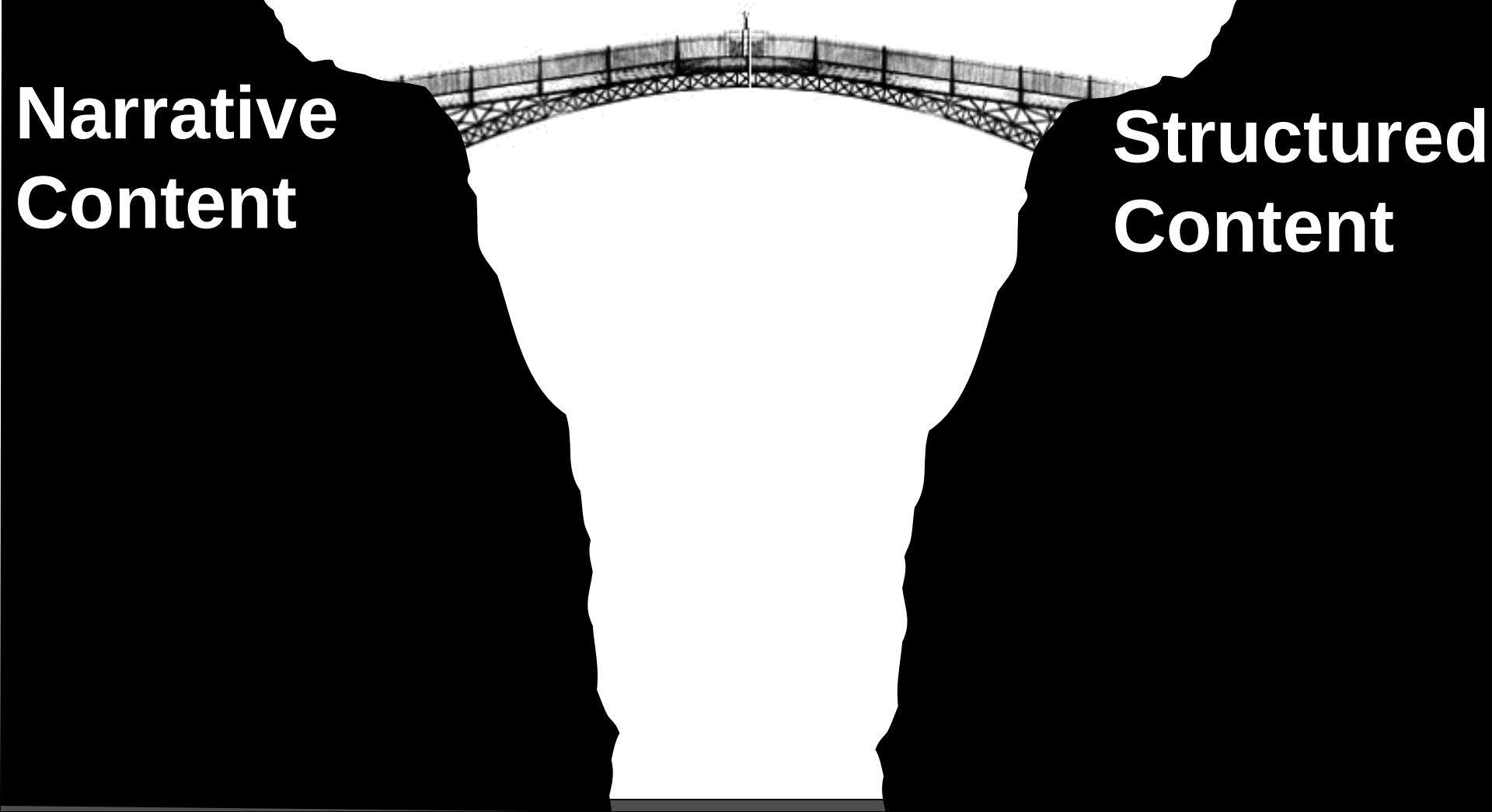
Narrative Content

finding reports
progress notes
discharge summaries

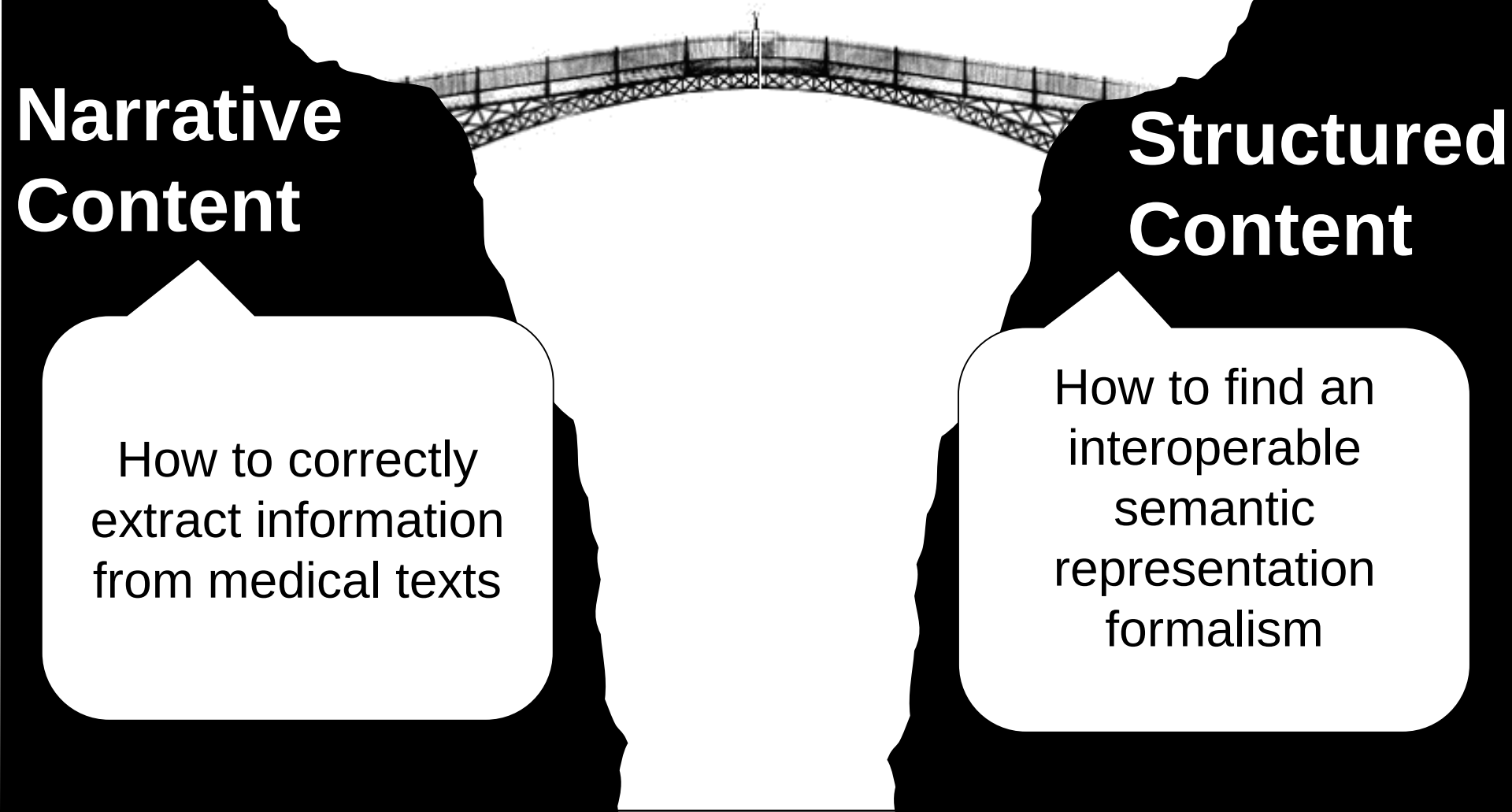
Structured Content

terminologies
classifications
ontologies
information models

Semantic enrichment of text using natural language technologies



Semantic enrichment of text using Natural Language Technologies



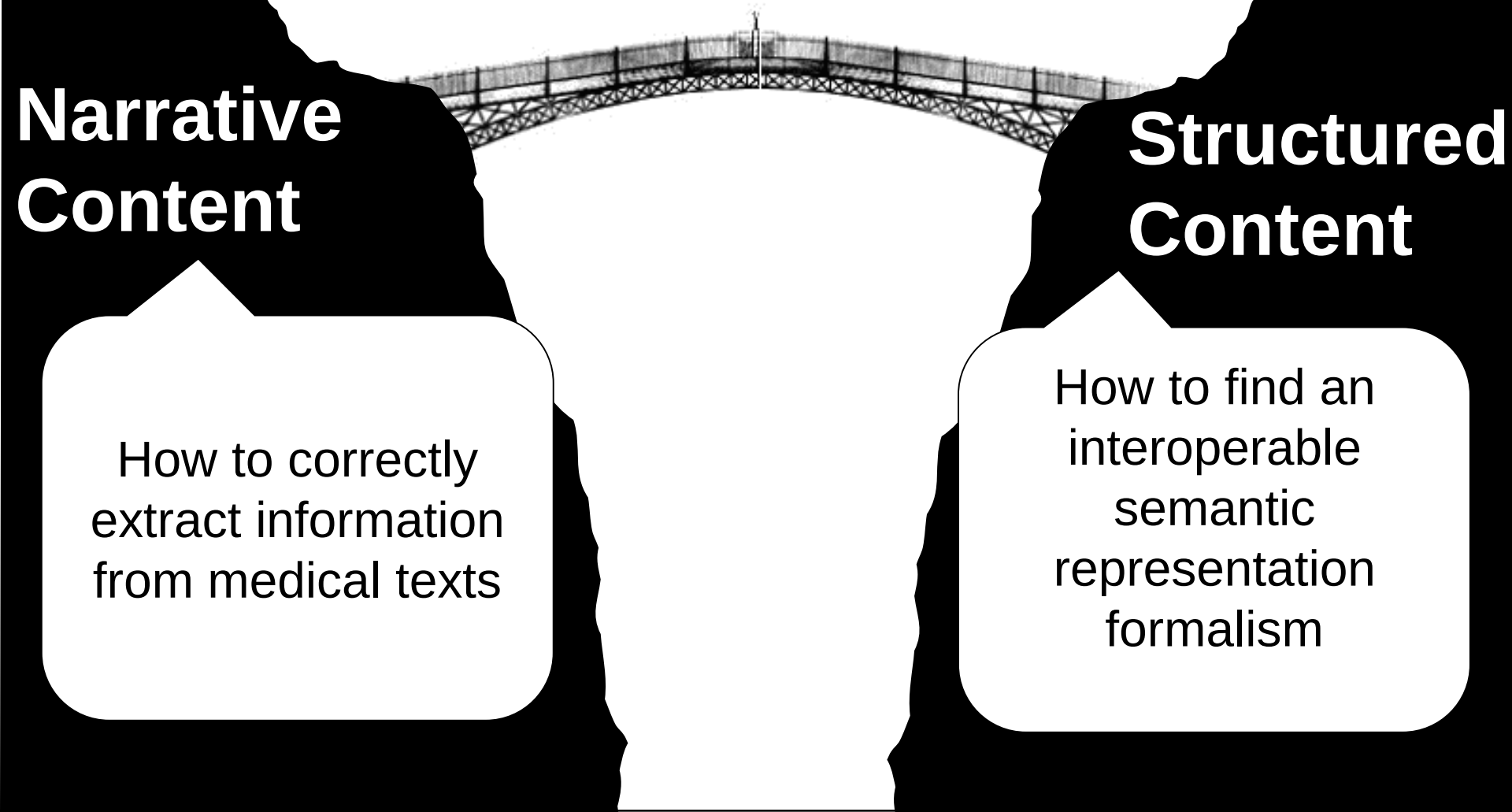
Narrative Content

How to correctly extract information from medical texts

Structured Content

How to find an interoperable semantic representation formalism

Semantic enrichment of text using Natural Language Technologies



Narrative Content

How to correctly extract information from medical texts

Structured Content

How to find an interoperable semantic representation formalism

Semantic enrichment of text using Natural Language Technologies

Narrative Content

How to correctly extract information from medical texts

Structured Content

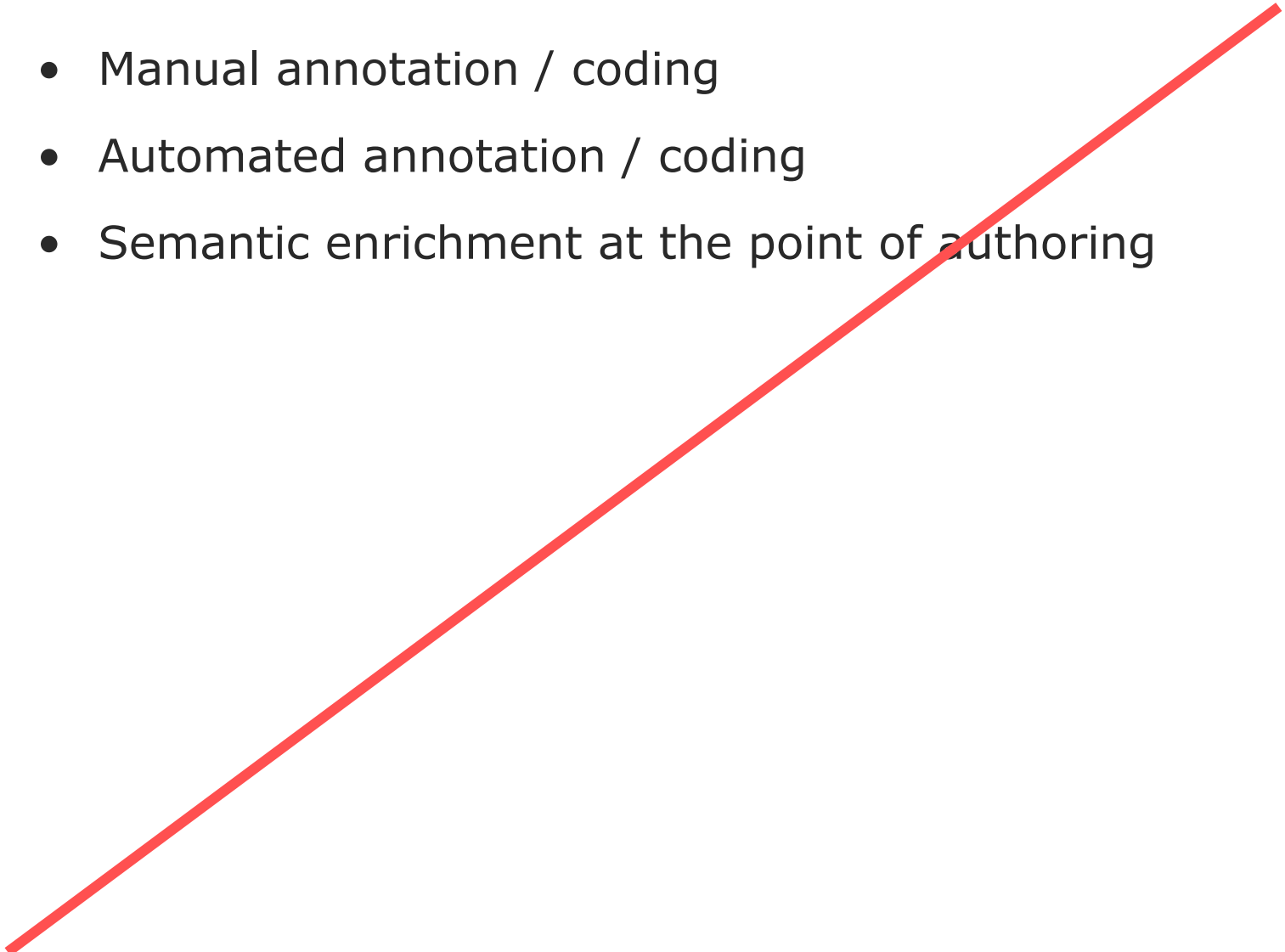
How to find an interoperable semantic representation formalism

Structure of the talk

- The importance of natural language in the EHR
- Semantic enrichment: Target representations
- Semantic enrichment: Technical challenges

Semantic enrichment: technical challenges

- Manual annotation / coding
- Automated annotation / coding
- Semantic enrichment at the point of authoring

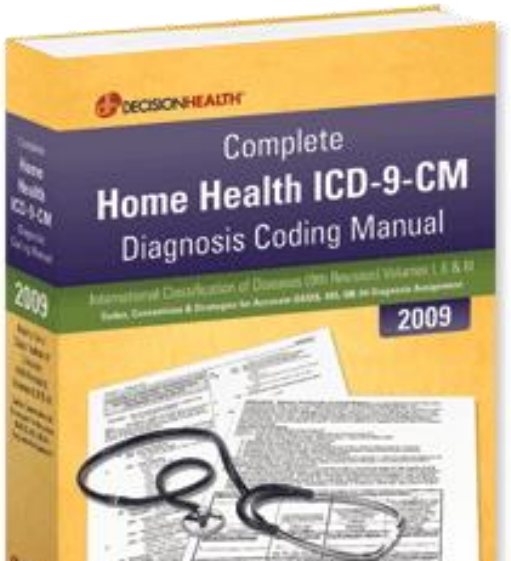


From narrative to structured content

- Semantic enrichment
  - Annotating text (or other data objects) by in ways that boost the value of the text
  - Semantic identifiers
    - ad-hoc tags
    - controlled terms
    - classes from ontologies and information models
- Strategies
  - Manual annotation / coding
  - Automated annotation / coding
  - Semantic enrichment at the point of authoring

Manual semantic enrichment

- Examples
  - MeSH indexing of Medline abstracts
  - Manual assignment of disease and procedure codes to the EHR

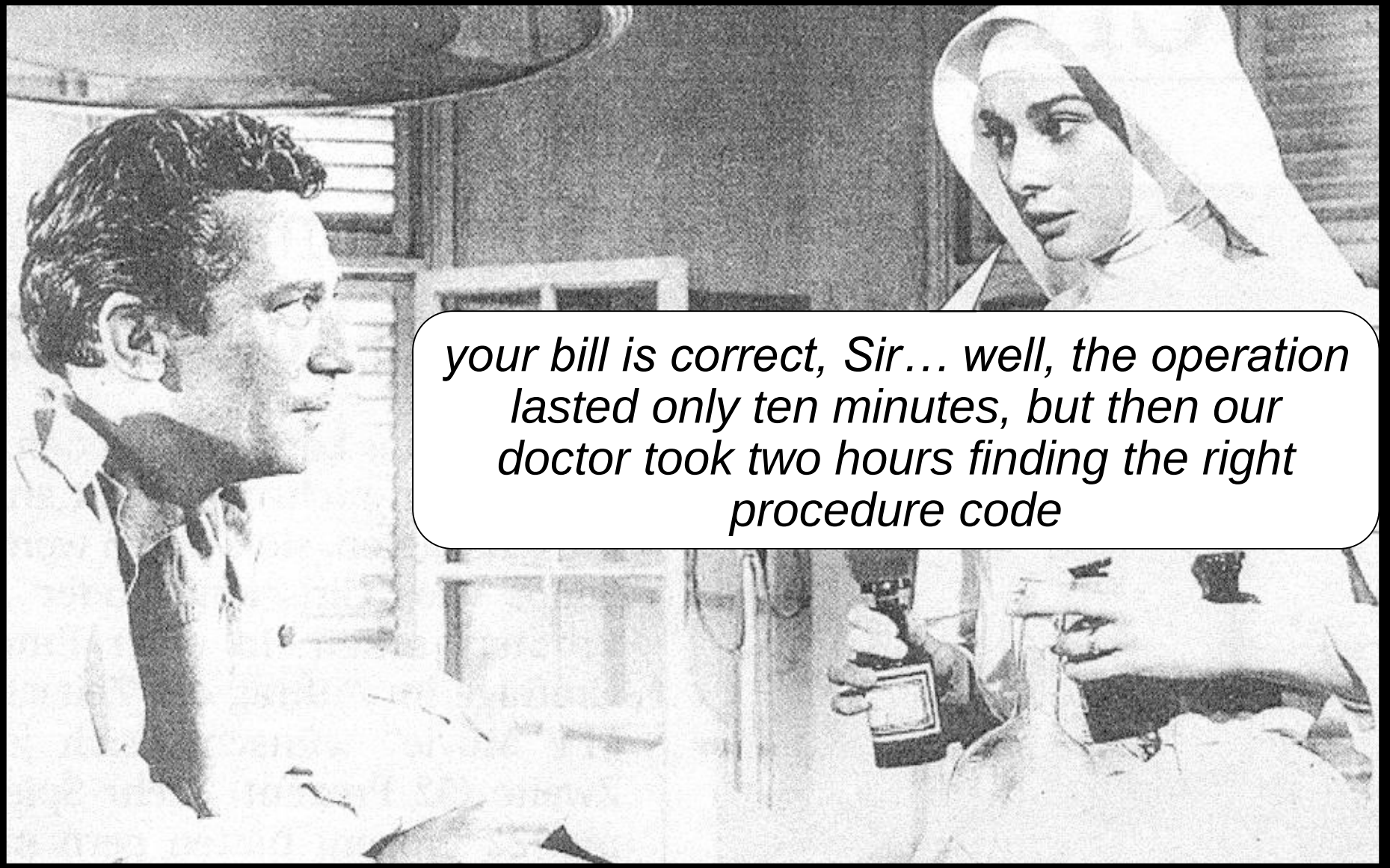


PubMed

MH	-	Adenovirus Infections, Human/epidemiology/virology
MH	-	Adenoviruses, Human/classification/isolation & purification
MH	-	Animals
MH	-	Birds
MH	-	Bocavirus/isolation & purification
MH	-	Bronchiolitis/diagnosis/*virology
MH	-	Communicable Diseases, Emerging/epidemiology/*virology
MH	-	Coronavirus/classification/isolation & purification
MH	-	Coronavirus Infections/epidemiology/virology
MH	-	Humans
MH	-	Infant
MH	-	Influenza A virus/classification/isolation & purification
MH	-	Influenza in Birds/epidemiology/virology
MH	-	Influenza, Human/epidemiology/virology
MH	-	Male
MH	-	Metapneumovirus/isolation & purification/*pathogenicity/physiology
MH	-	Mucocutaneous Lymph Node Syndrome/virology
MH	-	Paramyxoviridae Infections/diagnosis/epidemiology/*virology
MH	-	Parvoviridae Infections/epidemiology/virology
MH	-	Respiratory Tract Infections/epidemiology/*virology
MH	-	SARS Virus/isolation & purification
MH	-	Severe Acute Respiratory Syndrome/epidemiology/virology
MH	-	World Health
MH	-	Zoonoses9

Manual semantic enrichment

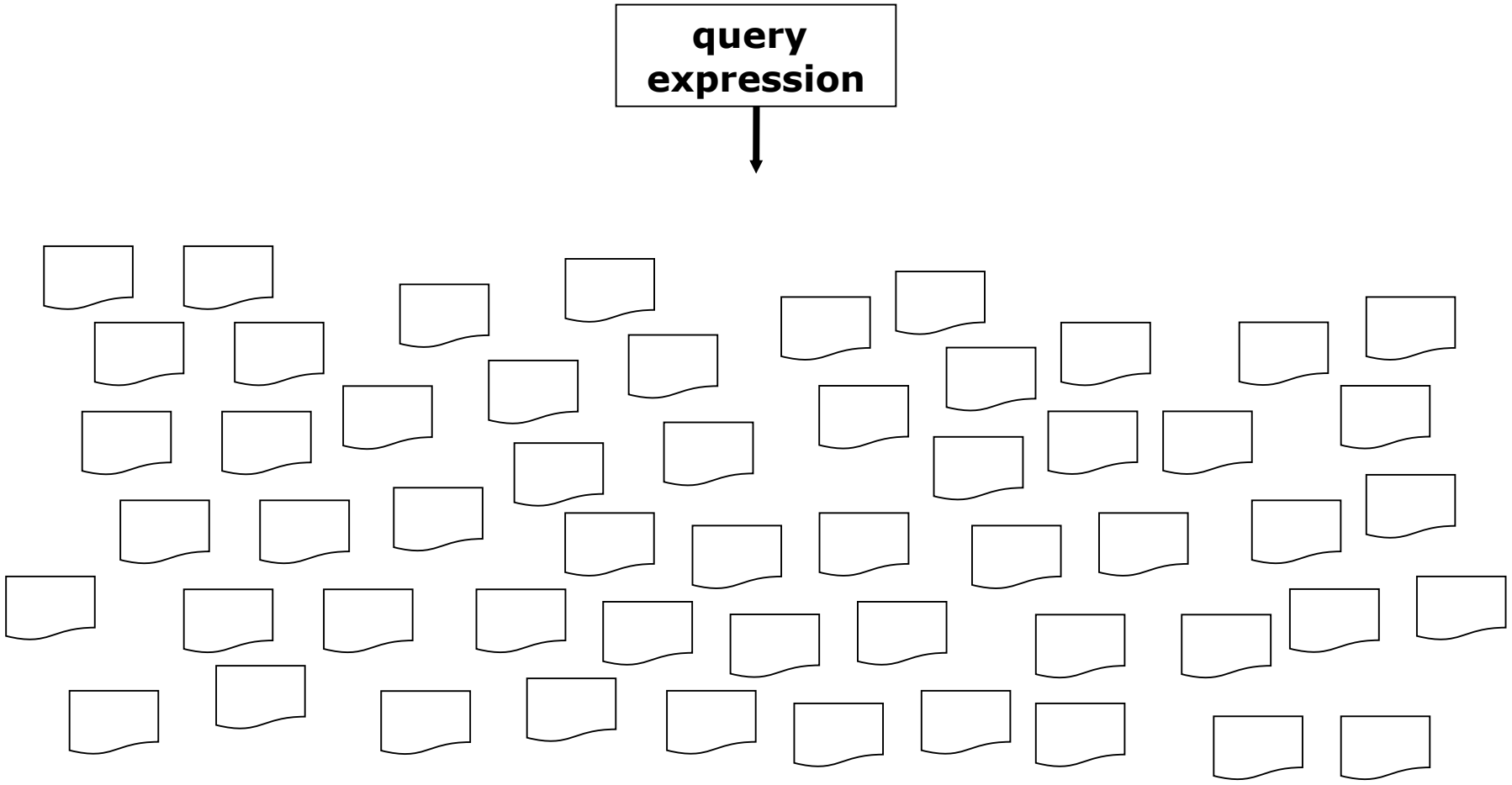
- Problems
  - Time consuming
  - Requires specific training
  - Motivation gap / bias:
    - undercoding
    - overcoding
    - miscoding
  - limited scope:
    - procedures
    - diseases


your bill is correct, Sir… well, the operation lasted only ten minutes, but then our doctor took two hours finding the right procedure code

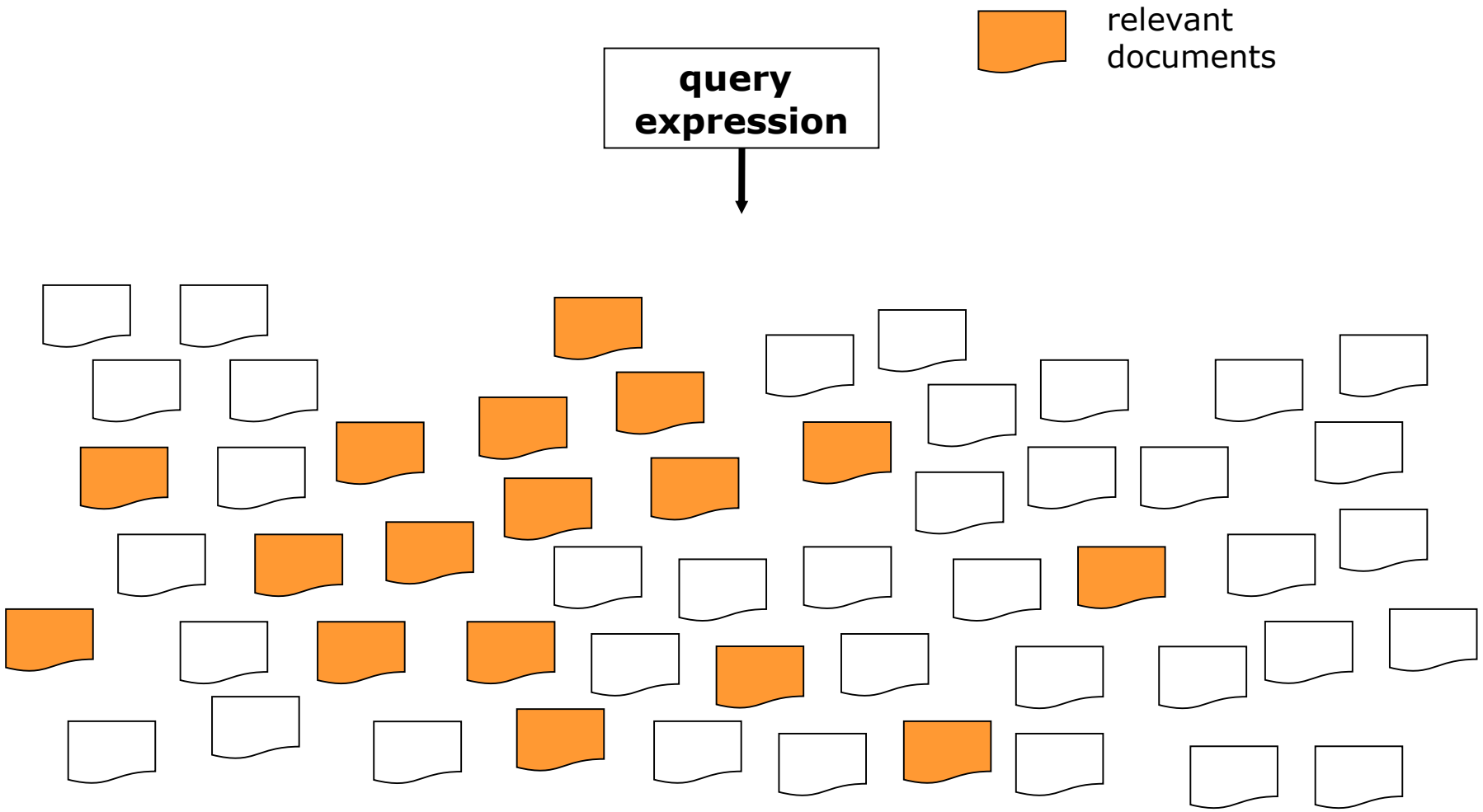
Automated semantic enrichment

- Uses natural language processing technology
- commonly used term: text mining
- Two paradigms
  - document retrieval
    special case: term retrieval
  - information extraction

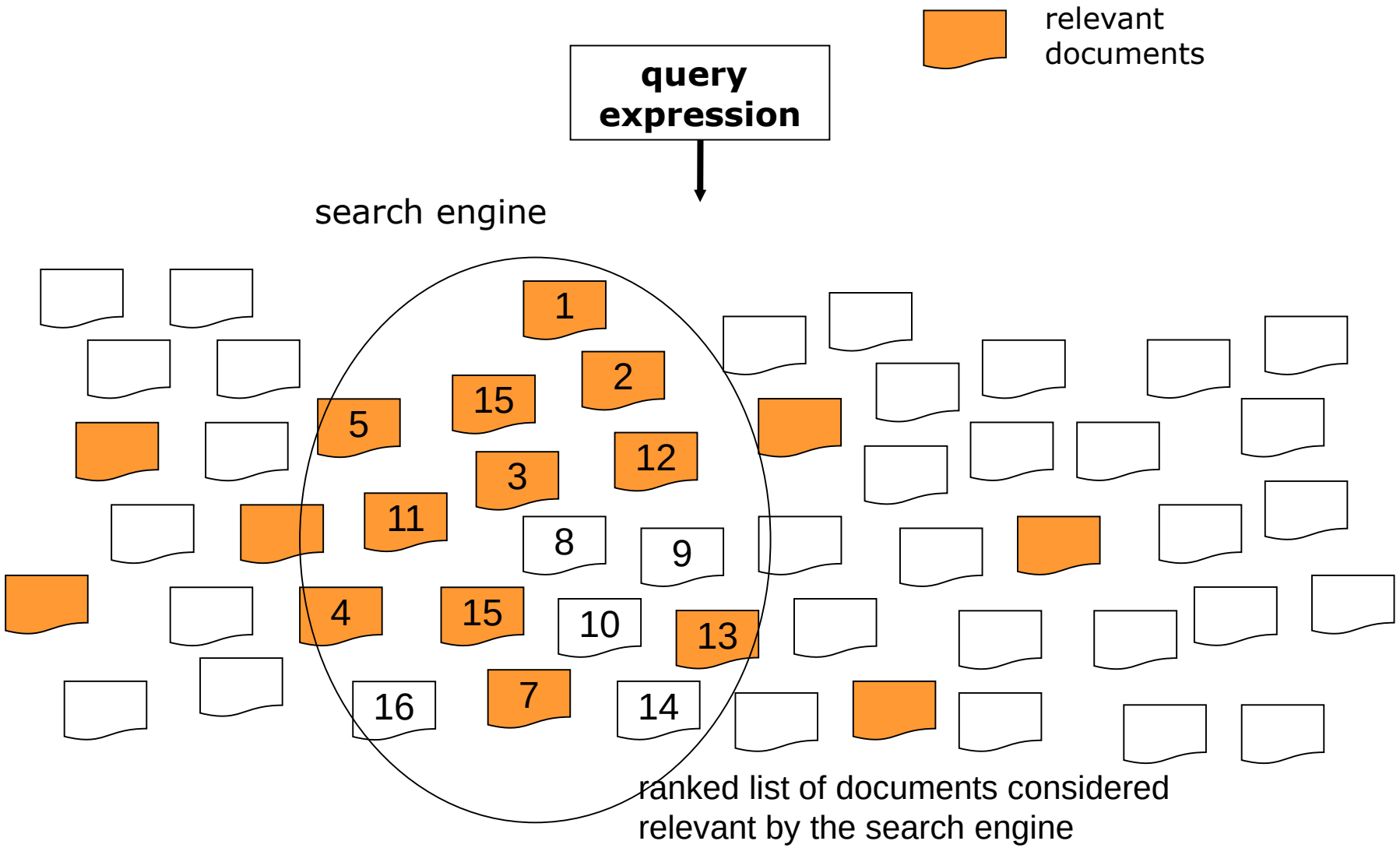
Document retrieval scenario



Document retrieval scenario

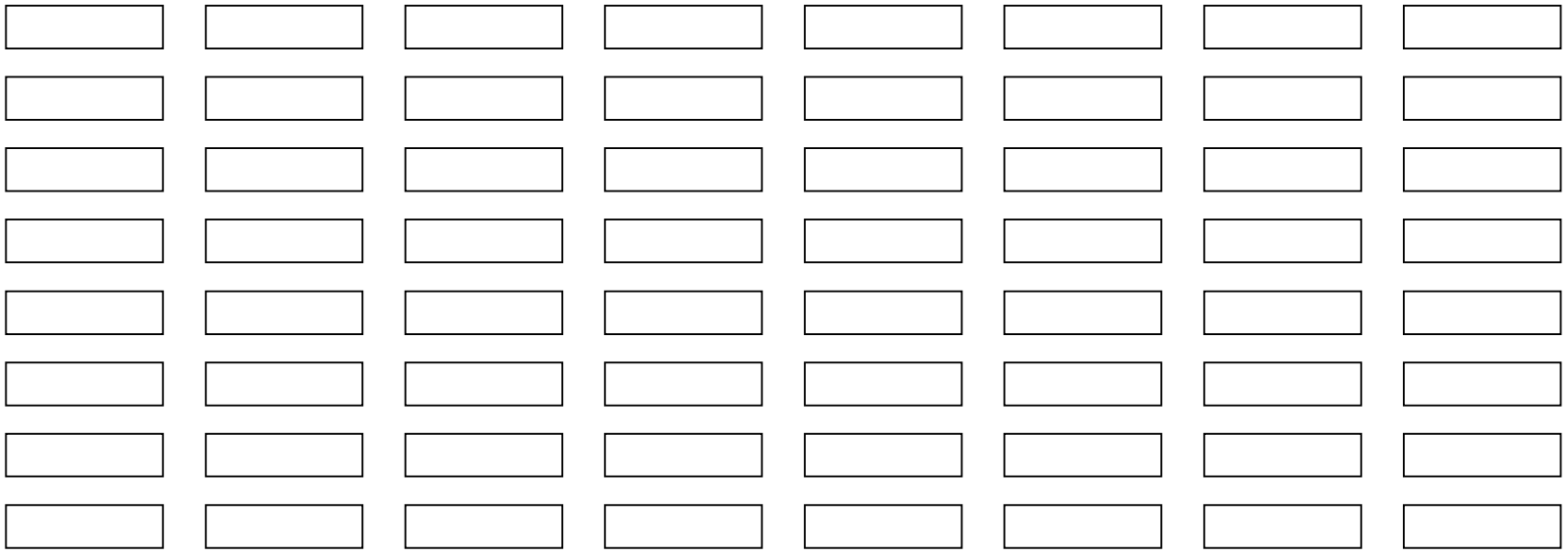


Document retrieval scenario



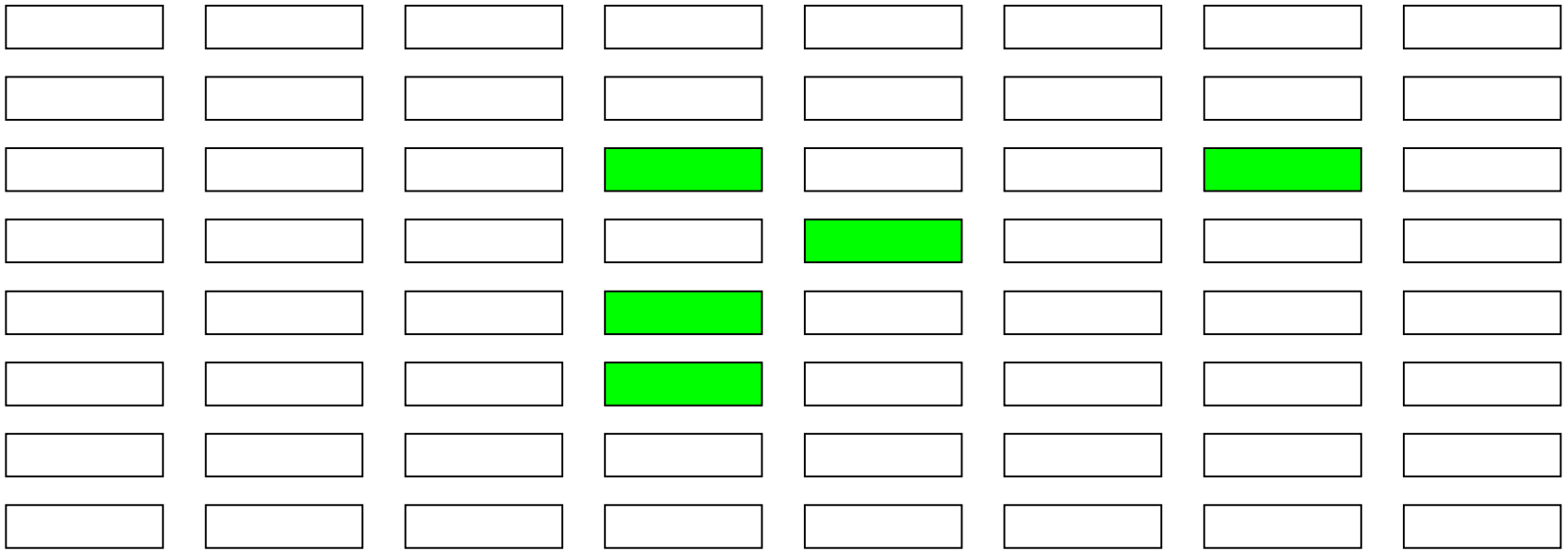
Term retrieval as a special case of document retrieval

AP: vasculite leucocitoclastica. A pcte recebeu ciprofloxacin por 5 dias e após 2 dias de oxacilina. Recebe alta em bom estado geral, com diminuição importante do edema e da dor em MID. Lesões em fase de cicatrização. Revisada a literatura: existe associação de vasculite em areas de estase, bem como associação com



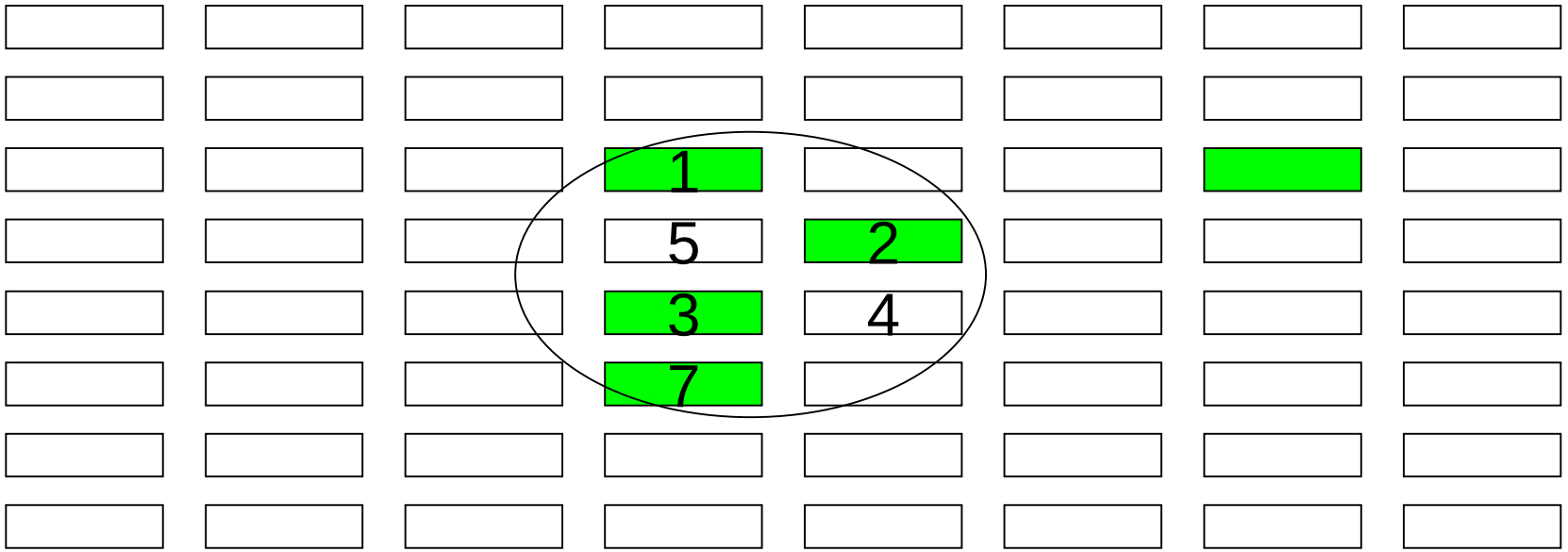
Term retrieval scenario

AP: vasculite leucocitoclastica. A pcte recebeu ciprofloxacin por 5 dias e após 2 dias de oxacilina. Recebe alta em bom estado geral, com diminuição importante do edema e da dor em MID. Lesões em fase de cicatrização. Revisada a literatura: existe associação de vasculite em areas de estase, bem como associação com



Term retrieval scenario

AP: vasculite leucocitoclastica. A pcte recebeu ciprofloxacin por 5 dias e após 2 dias de oxacilina. Recebe alta em bom estado geral, com diminuição importante do edema e da dor em MID. Lesões em fase de cicatrização. Revisada a literatura: existe associação de vasculite em areas de estase, bem como associação com



ranked list of terms considered
matching candidates by the search engine

Information extraction: example

shadow was pointed out on a routine chest X-ray film, but she had no further examination. Physical examination on admission revealed purpura of the upper and lower extremities, swelling of the gums and tonsils, but no symptoms showing the complication of myasthenia gravis. Hematological tests revealed leucocytosis: WBC count 68 700/μl (blasts 11.5%, myelocytes 0.5%, bands 2.0%, segments 16.0%, monocytes 65.5%, lymphocytes 4.0%, atypical lymphocytes 0.5%), Hb 7.1 g/dl (reticulocytes 12%) and a platelet count of 9.1×10^4/μl. Further laboratory examination revealed elevated serum lactic dehydrogenase (589 U/l), vitamin B_{12} (2010 pg/ml) and ferritin (650. 0 ng/ml). Human chorionic gonadotropin and [alpha]-fetoprotein levels were normal. A bone marrow aspiration revealed hypercelllar bone marrow with a decreased number of erythroblasts and megakaryocytes and an increased number of monoblasts that were positive for staining by [alpha]-naphthyl butyrate esterase and negative for staining by naphthol ASD chloroacetate esterase. Chest X-ray upon admission revealed a mediastinal mass and an elevated left diaphragm . Computed tomography (CT) of the chest showed a left anterior mediastinal mass . Based on these findings, the patient was diagnosed with a mediastinal tumor accompanied by AMoL. First, in June 1991, the patient was treated with DCMP therapy: daunorubicin (DNR) (25 mg/m^2, days 1, 2, 3, 4, 6 and 8), cytosine arabinoside (Ara-C) (100 mg/m^2, days 1-9), 6-MP riboside (6-MP) (70 mg/m^2, days 1-9) and prednisolone (PSL) (20 mg/m^2, days 1-9), followed by five courses of consolidation chemotherapy [1, DCMP; 2, ID-Ara-C:adriacin (ADR), vincristine (VCR), Ara-C, PSL; 3, DCMP; 4, ID-Ara-C; 5, A-triple V: Ara-C, VP-16, VCR, vinblastine (VBL)]. After induction chemotherapy, a hematological examination and bone marrow findings had improved to normal, and complete remission was attained. Chest CT scan after chemotherapy in November 1991 revealed regression of the mediastinal tumor . An invasive thymic tumor was suspected and surgery was undertaken in January 1992. The tumor (50 × 45 × 45 mm), located mainly in the anterior mediastinum, was strongly adhered to the adjacent tissues. Resection of the tumor included the left upper lobe of the lung, the phrenic nerve and pericardium. The histological finding was that the tumor cells have large, vesicular nuclei and prominent nucleoli, but keratinazation was unclear . The results of immunohistochemical finding of anti-TdT was negative. From these findings, we diagnosed poorly or moderately differentiated squamous cell carcinoma of the thymus. The postoperative course was uneventful. The patient underwent radiation therapy of the mediastinum and left hilum at doses of 4000 cGy delivered over 4 weeks. She was discharged in March 1992. After the first AMoL remission, the patient suffered a relapse six times and was repeatedly admitted for chemotherapy. During these periods, chest X-ray and CT revealed no recurrence of the mediastinal tumor. During her tenth admission, the patient developed pneumonia during chemotherapy and died in October 1996. No autopsy was performed.

Tumor registry - template	
date of first diagnosis	
primary loca-lization	
grading	
staging	
morphology	
Date primary therapy	
chemotherapy	
radiation	

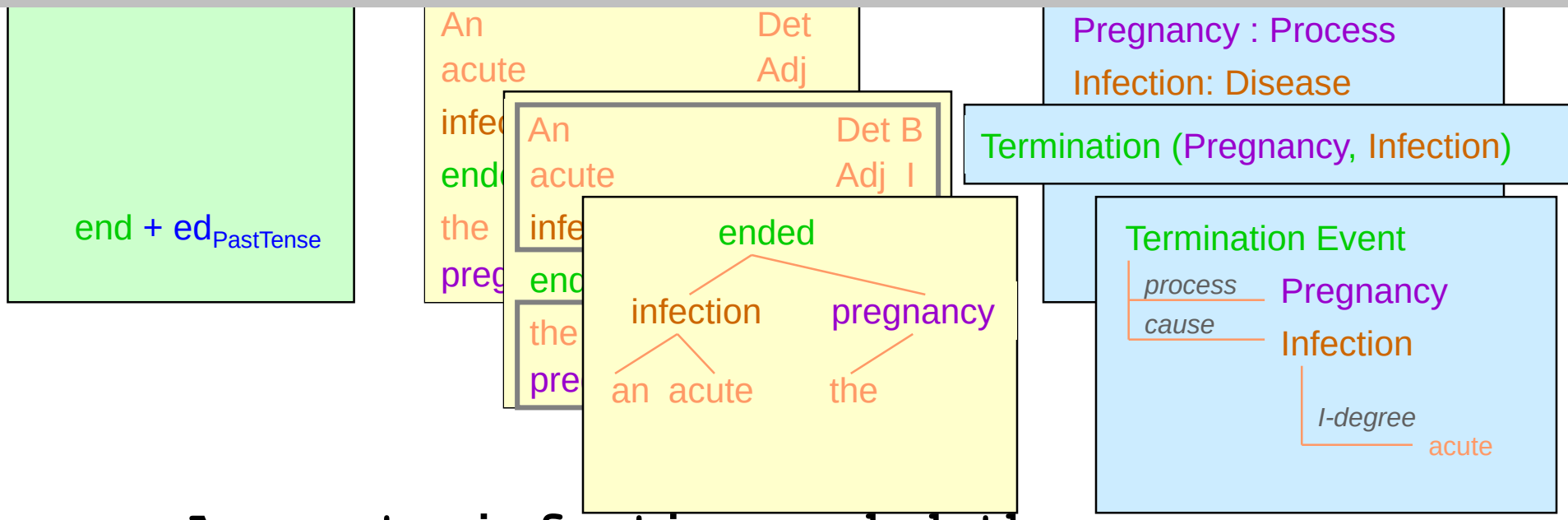
Language processing techniques

- from simple pattern matching...

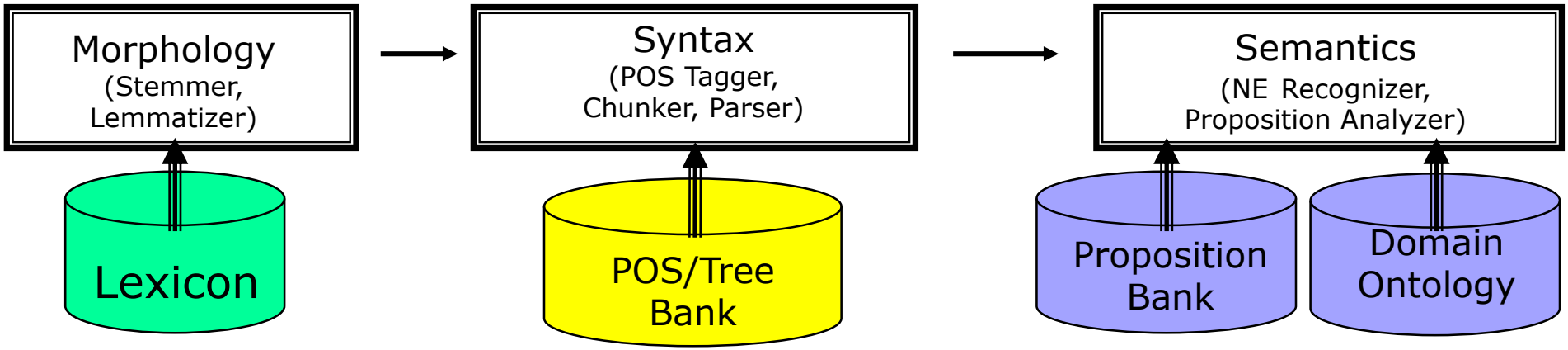
  "do*" → "do", "dog", "done", "doctor", etc.

- to more sophisticated techniques, using
  - lexical knowledge
  - grammatical knowledge
  - domain knowledge
  - empirical knowledge (e.g. **annotated corpora**)
  - AI approaches
  - statistical approaches

Pipeline for NLP Analysis



«An acute infection ended the pregnancy»



example: Udo Hahn, Jena

Possible outputs of sophisticated language processing

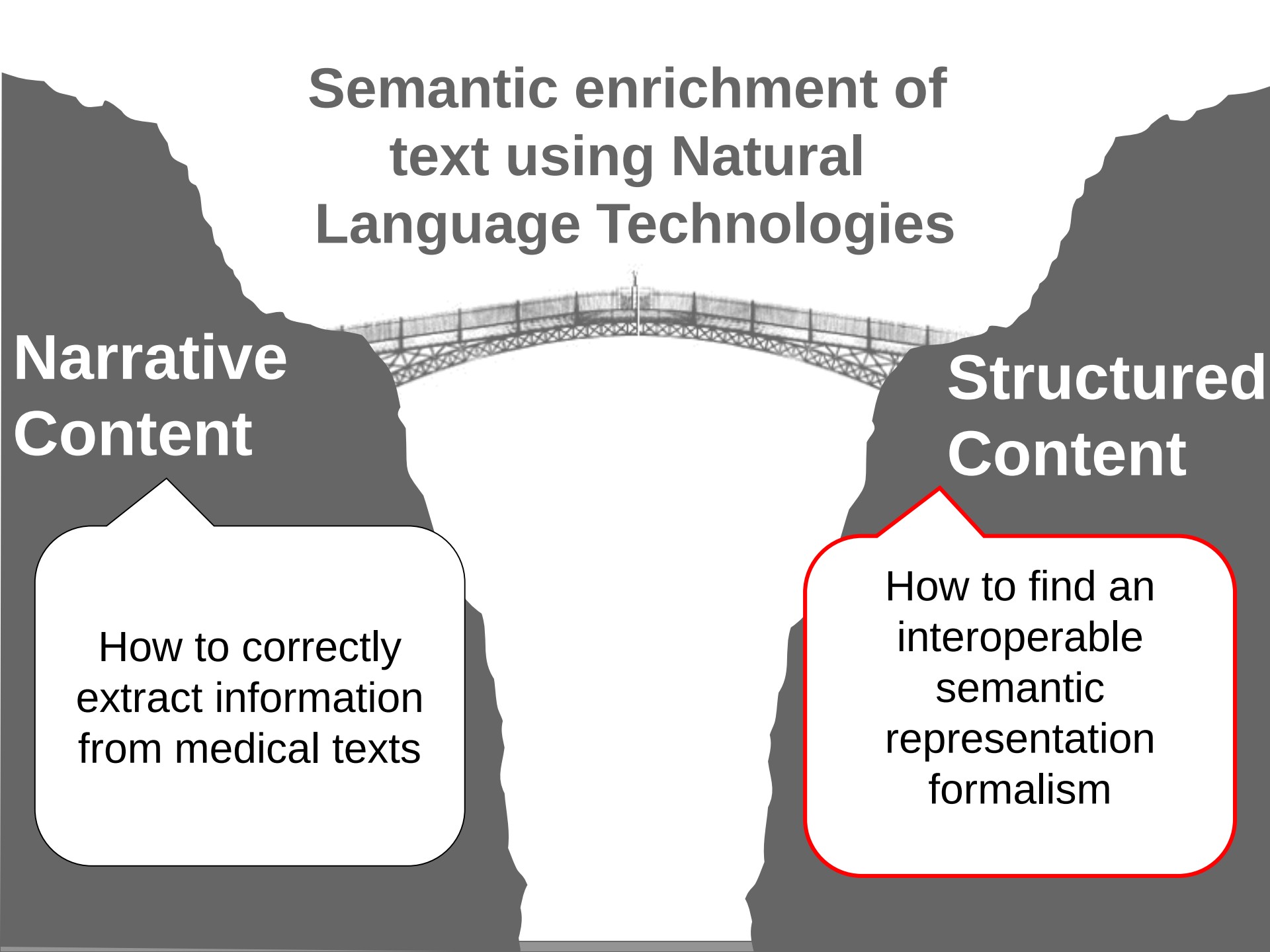
- Noun phrases (term candidates)
- Predicate-argument structures
- classification of named entities
- attachment of prepositional phrases
- scope of negations
- anaphora resolution
- discourse analysis
- etc…

text “understanding” is still the holy grail of computational linguistics

Specific challenges for medical language processing

- High lexical productivity
  - single-word compounds “hyperparathyroidism”
  - acronyms and abbreviations, ambiguous and context-dependent
- heterogenous document style
  - telegram style
  - enumerations (e.g. lab values)
  - embedded tables
- low writing quality
  - persisting errors (spelling, punctuation, case, accents…)
- implicit contexts

Semantic enrichment of text using Natural Language Technologies



Narrative Content

How to correctly extract information from medical texts

Structured Content

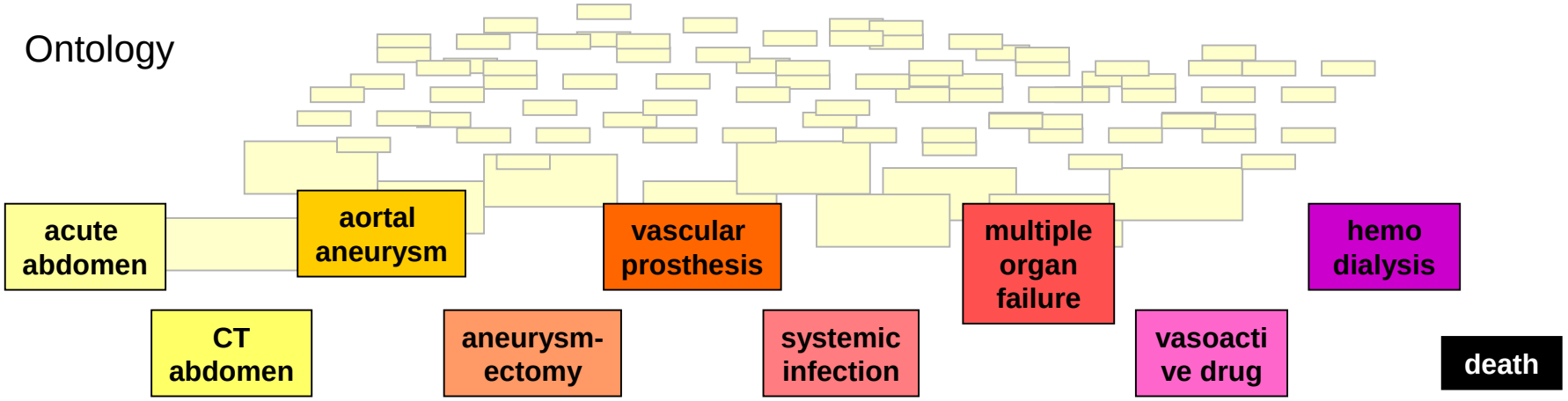
How to find an interoperable semantic representation formalism

Naïve approach: content representation by instantiation of ontologies

- Identify term of interest *T* in a text
- Retrieve a suitable class in the ontology using term retrieval
- Interpret the mention of this term as the reference to a member of this class
- i.e.

  T mentioned in text = there is some referent of *T* in reality

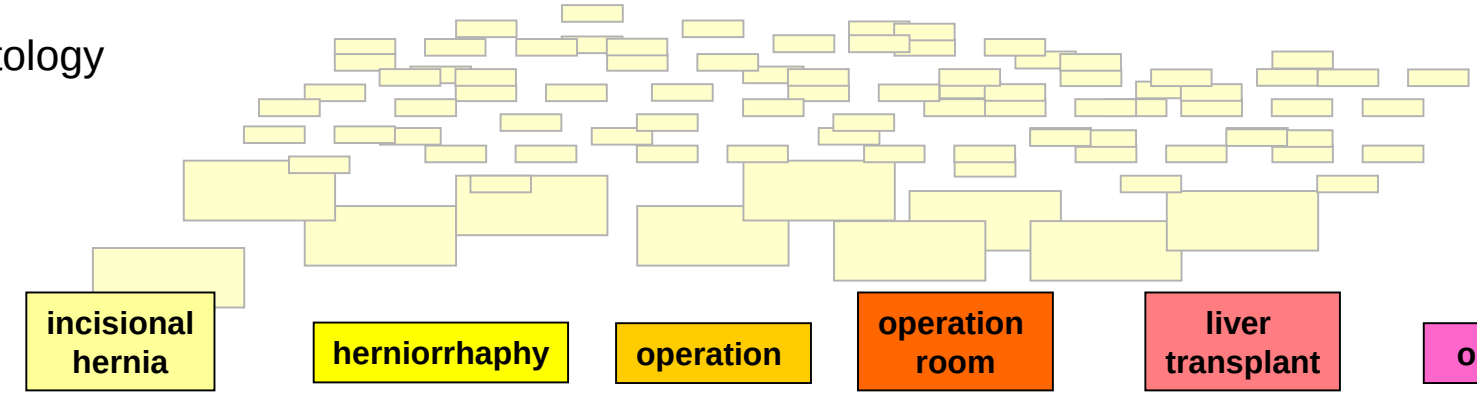
Example



Paciente interna por quadro de abdome agudo. TC de abdome mostrou aneurisma de aorta com evidencia de extravasamento de contraste. Levado a cirurgia de urgência, sendo realizada aneurismectomia com colocação de prótese. No pós-operatório evoluiu com síndrome da resposta inflamatória sistêmica, com disfunção de múltiplos órgãos e instabilidade hemodinâmica. Apesar do manejo com drogas vasoativas, reposição hídrica e hemodiálise veno-venosa crônica, o paciente apresentou piora progressiva, evoluindo para óbito

Counterexample (more realistic)

Ontology



- instance

- not instantiated
- reference to a plan

- reference to a suspended plan

- instance

- reference to a plan
- NOT referring to the same patient

- reference to a new plan
- not yet executed

Patient with incisional hernia admitted for herniorrhaphy, but operation was suspended because operation room was urgently needed for liver transplant. Discharged with orientation and rescheduled operation

Example: codes and contexts

Chunk	Context	Negation	Code (SNOMED)
O cateter			19923001
foi trocado			103713001
por disfunção			
(baixo fluxo)			
no mesmo sítio.			
Em discussão	HYP		
papilotomia ou	HYP		235582004
colecistectomia	HYP		38102005
com a Gastro,	HYP		71838004
mas como não tem			
cálculo		NEG	56381008
essa decisão			
será tomada posteriormente.			
Em 31/01/07			
apresentou			
PCR			410430005
em fibrilação ventricular			71908006
logo após			
ter terminado			
a hemodiálise;	STP		302497006
recuperado rapidamente			
não ficando			
com sequelas.		NEG	362977000

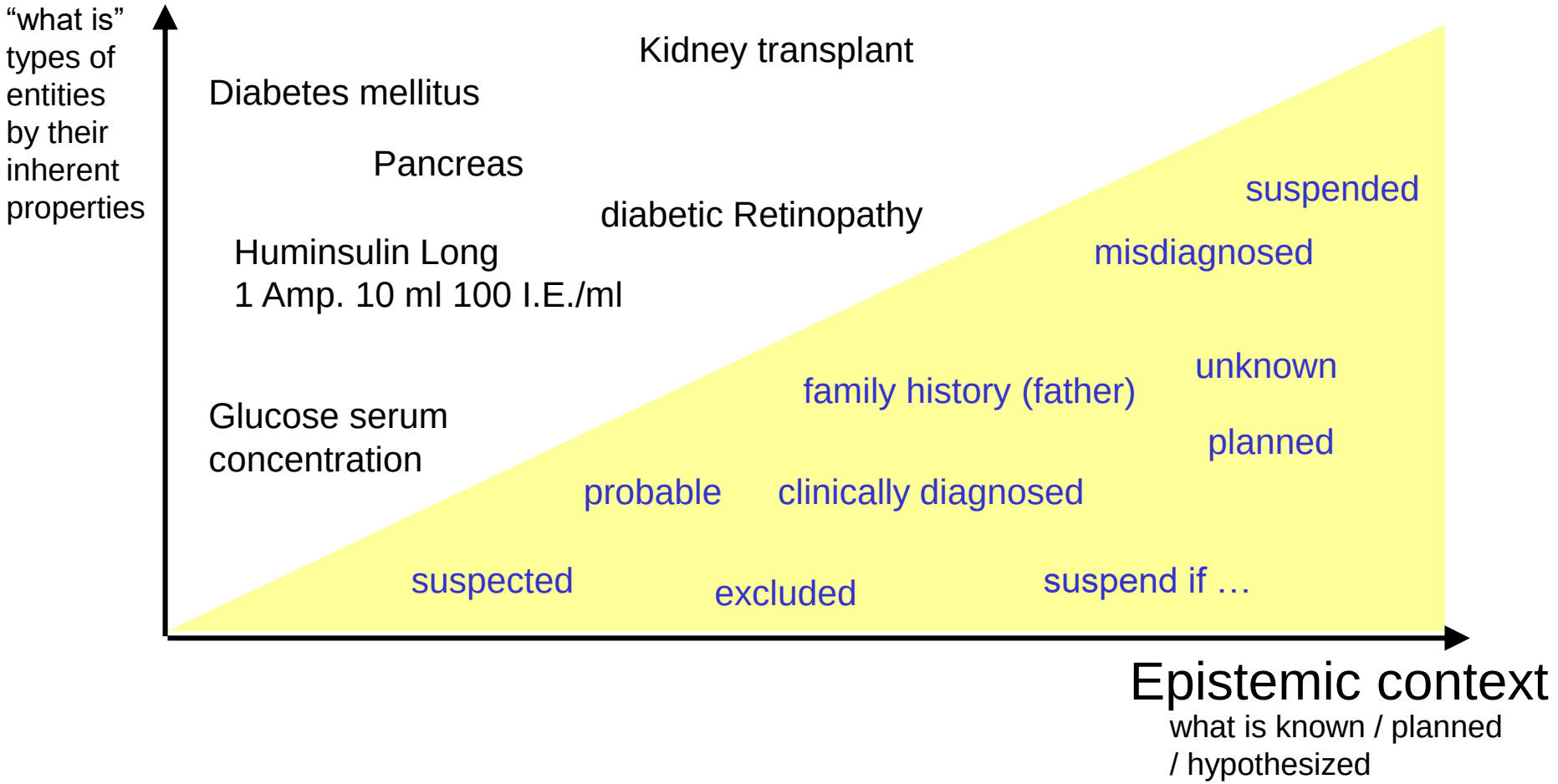
Why ontologies are not enough

Common contexts in medical documents:

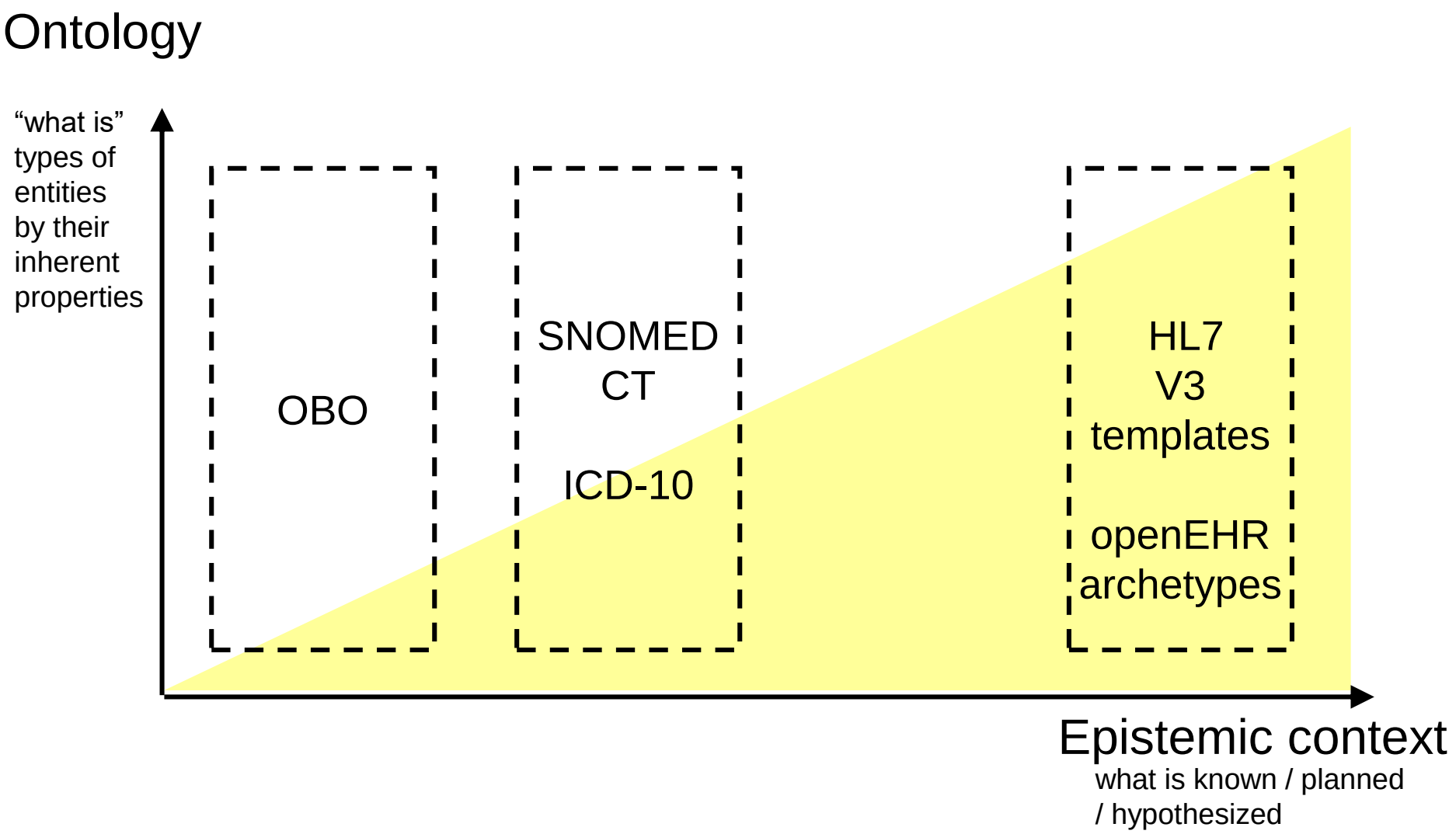
- Subject of record:
  - "father had diabetes mellitus"
- Uncertainty
  - "…was admitted with suspected diabetes mellitus"
- Negation
  - "no diabetes mellitus"
- Plan
  - "in the case of … patient should be checked for diabetes mellitus"

Boundary problem

Ontology

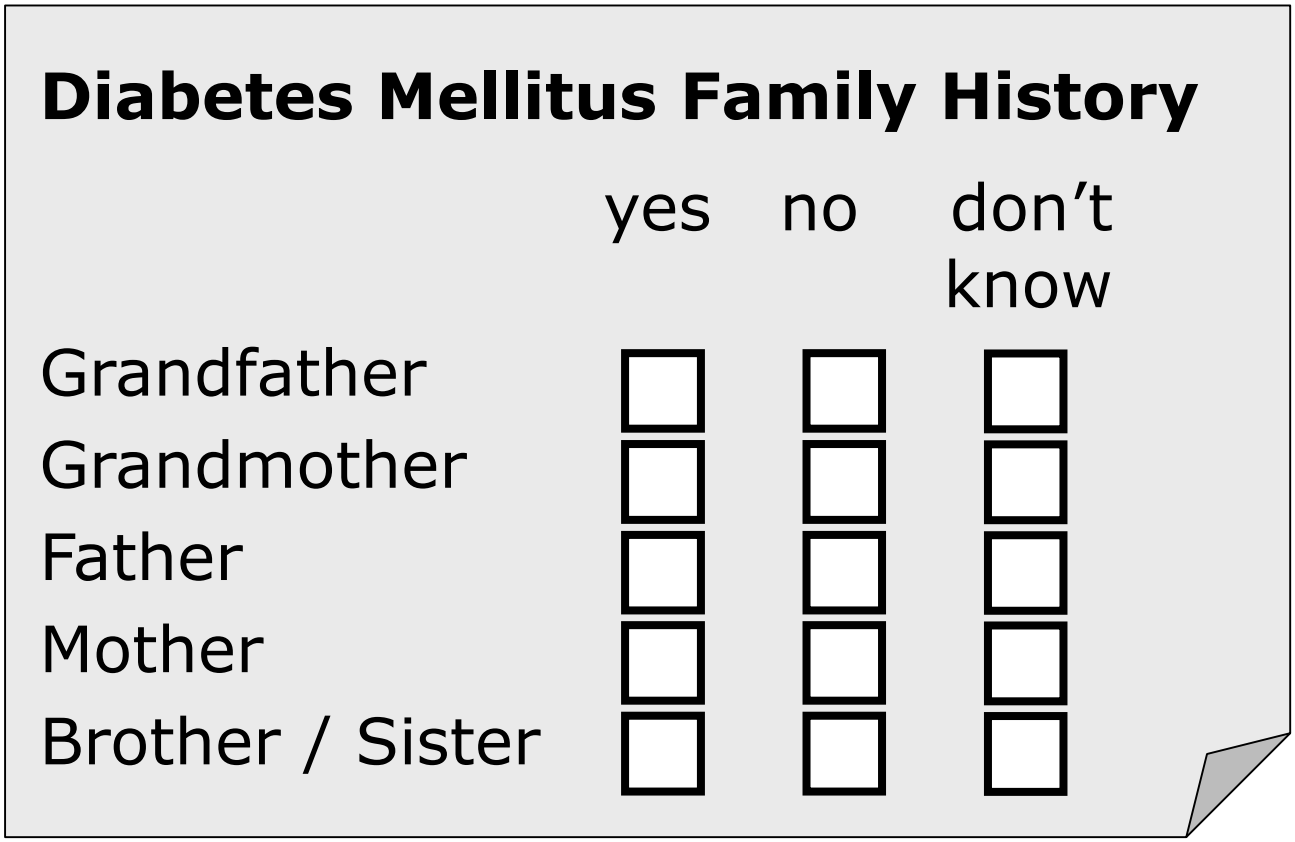


Ontology vs. Epistemology



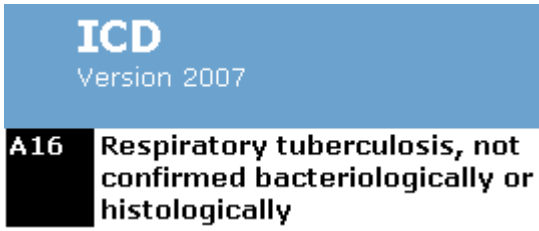
Information models

mix ontological with epistemic information



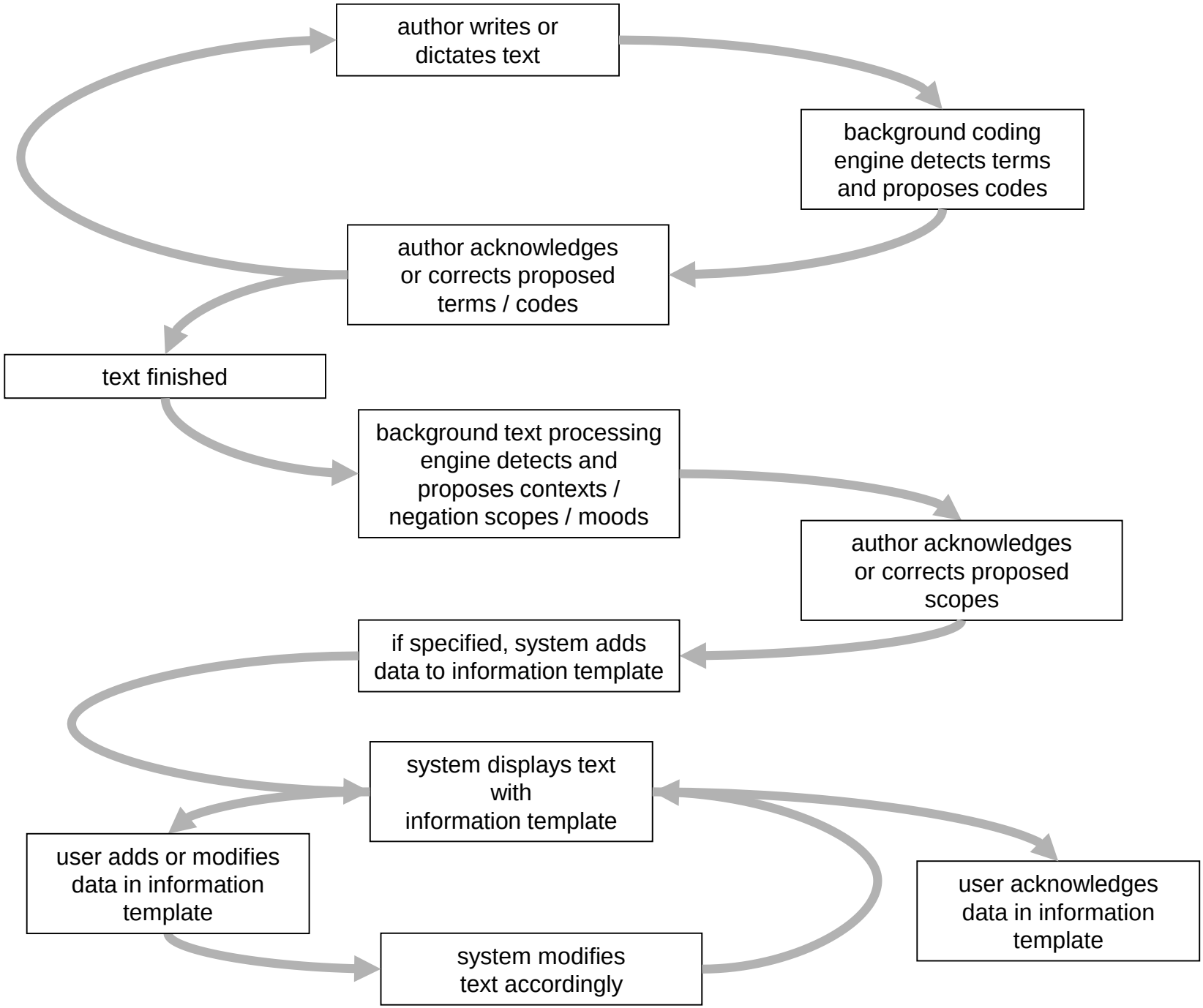
Conclusion for semantic representation of clinical narratives

- Clinical narratives combine
  - reference to types of entities (ontology)
  - reference to the state of knowledge of the author (context)
- Ignore context: high risk of false assertions
- Ignore ontology: limited semantic interoperability
- Binding ontologies to information models:
  - topic of research
  - pragmatic solutions
    (TermInfo: http://www.hl7.org/v3ballot/html/welcome/environment/index.htm)
- Using terminological systems that include both: current practice, but problematic



Outlook

- Persisting problem:
  - physicians continue producing text and coded content partly redundantly
- Challenge:
  - semantic enrichment on the fly


author writes or dictates text
background coding engine detects terms and proposes codes
author acknowledges or corrects proposed terms / codes
text finished
background text processing engine detects and proposes contexts / negation scopes / moods
author acknowledges or corrects proposed scopes
if specified, system adds data to information template
system displays text with information template
user adds or modifies data in information template
system modifies text accordingly
user acknowledges data in information template